확률

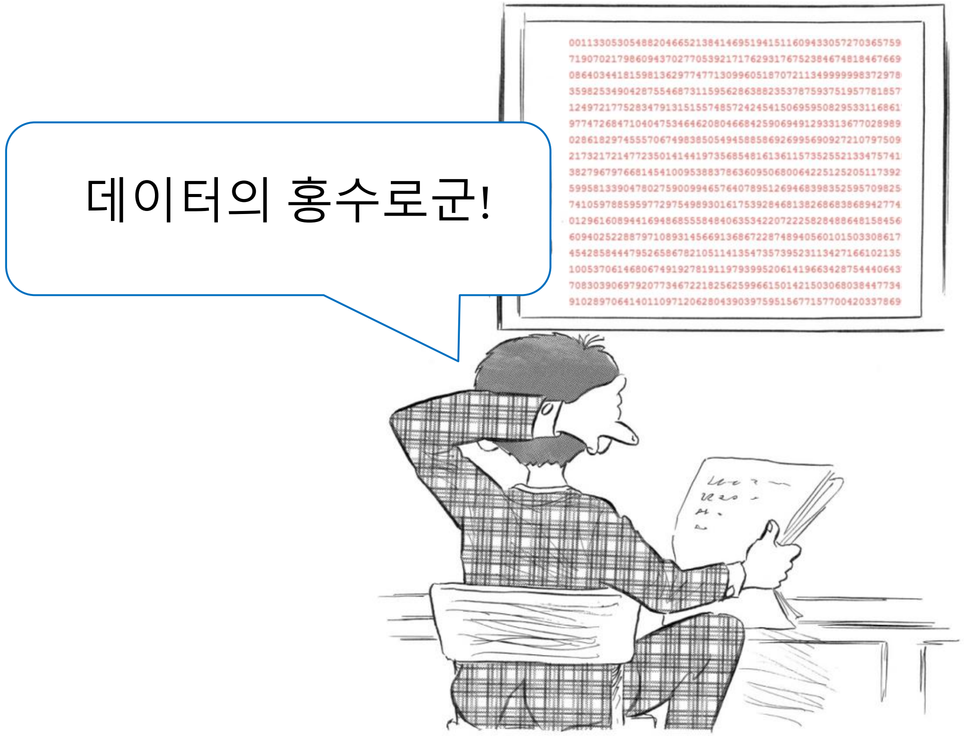
모듈 - 1

강사: 장순용 박사

순서

시대적 배경

지금은 소위 AI/빅데이터 시대라고 하는데....



시대적 배경

AI와 빅데이터는 이미 생활의 일부가 되어 있습니다:

실상

그런데 AI/빅데이터를 제대로 이해하고 있는 사람은 많지 않습니다.

장님이 코끼리 어루 만지듯…

큰 그림을 보아야 하는데…

주목해야할 트렌드

이세돌 대 알파고의 충격적인 결과:

- 이세돌 9단 대 알파고: 1승 4패.

주목해야할 트렌드

이세돌 대 알파고의 충격적인 결과:

- 이세돌 9단 대 알파고: 1승 4패.
- 인간의 고유 영역에 대한 AI의 거침없는 도전.
- 예상했던 것 보다 빠른 AI의 발전속도.
- 기업체(Google, Facebook, Amazon 등) 주도로 발전, 경제적 가치 창출!

주목해야할 트렌드

“MIT 디지털 경제 이니셔티브” 연구소장 Andrew McAfee:

- Youtube 동영상 “미래의 직업 환경”.

주목해야할 트렌드

“MIT 디지털 경제 이니셔티브” 연구소장 Andrew McAfee:

- Youtube 동영상 “미래의 직업 환경”.
- 스마트화, 자동화에 인한 일자리 감소.
- 급진적인 AI의 발전 때문에 ‘화이트 칼라’ 일자리도 줄어들고 있습니다.
- 오히려 블루 칼라 일자리의 감소세는 둔화되고 있습니다.
- 소득의 평균은 꾸준히 증가하지만 중위수는 정체되어 있습니다 → 양극화!

현재 진행중인 변화

사람 중심

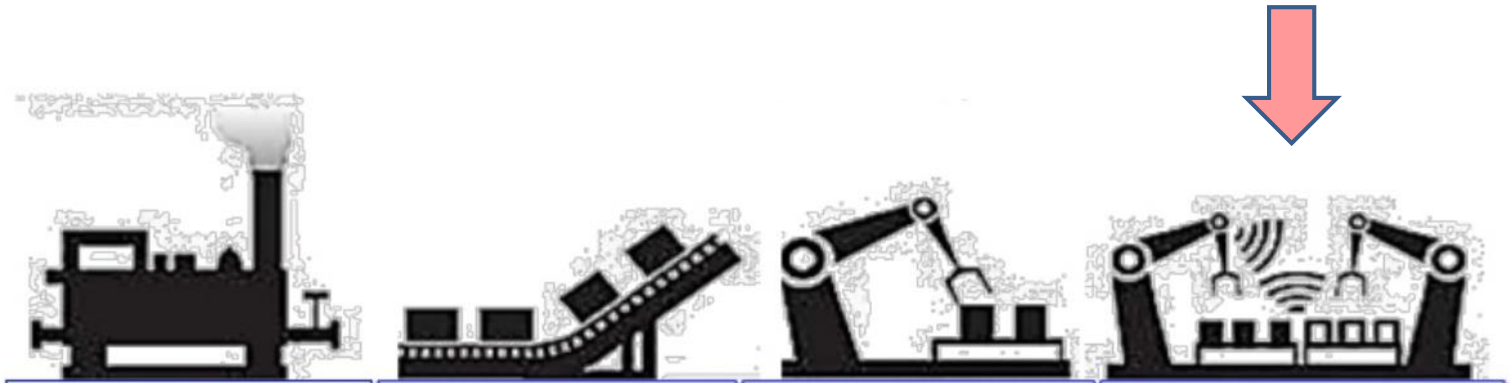
Machine 지원

Machine 주도

시간

- Machine 주도형의 장점: 빅데이터를 사용한 합리적이고 최적화된 판단 가능.
- 데이터가 사용되는 모든 산업 활동에 적용 가능 : 엄청난 파급효과.
- 인간의 인지력이 개입하는 곳에 새로운 기회가 있습니다 ← 준비된 인재.

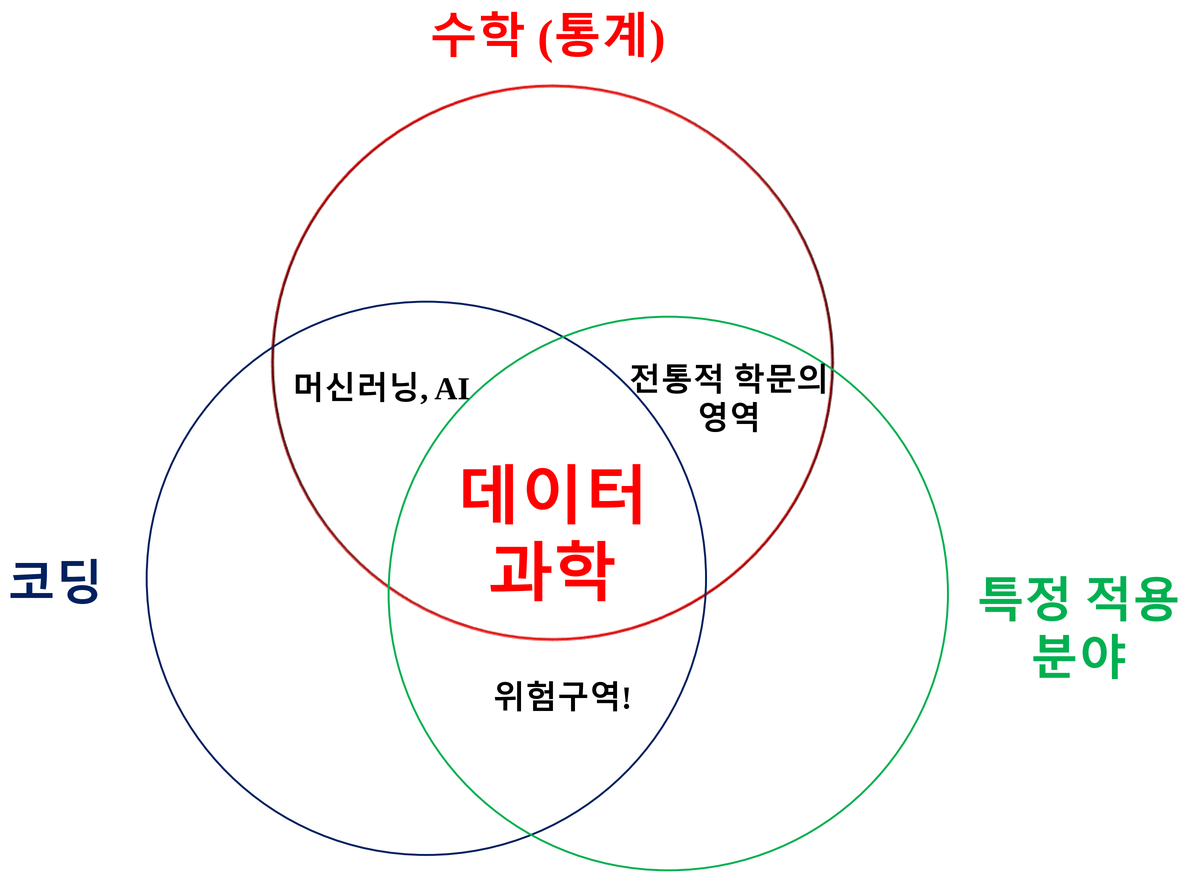
현재 진행중인 변화 : 4차 산업 혁명



1차 산업혁명	2차 산업혁명	3차 산업혁명	4차 산업혁명
18세기	20세기초	1970년	2015년
증기기관 발명	컨베이어벨트	공장 자동화	사물인터넷/AI
공장 생산체제	대량생산 체제	산업용 로봇	사이버물리시스템
기계화 혁명	대량생산 혁명	지식정보 혁명	만물초지능 혁명
공장 근로자	기술자	지식 근로자	융복합 창의인재

현재 진행중인 변화 : 4차 산업 혁명

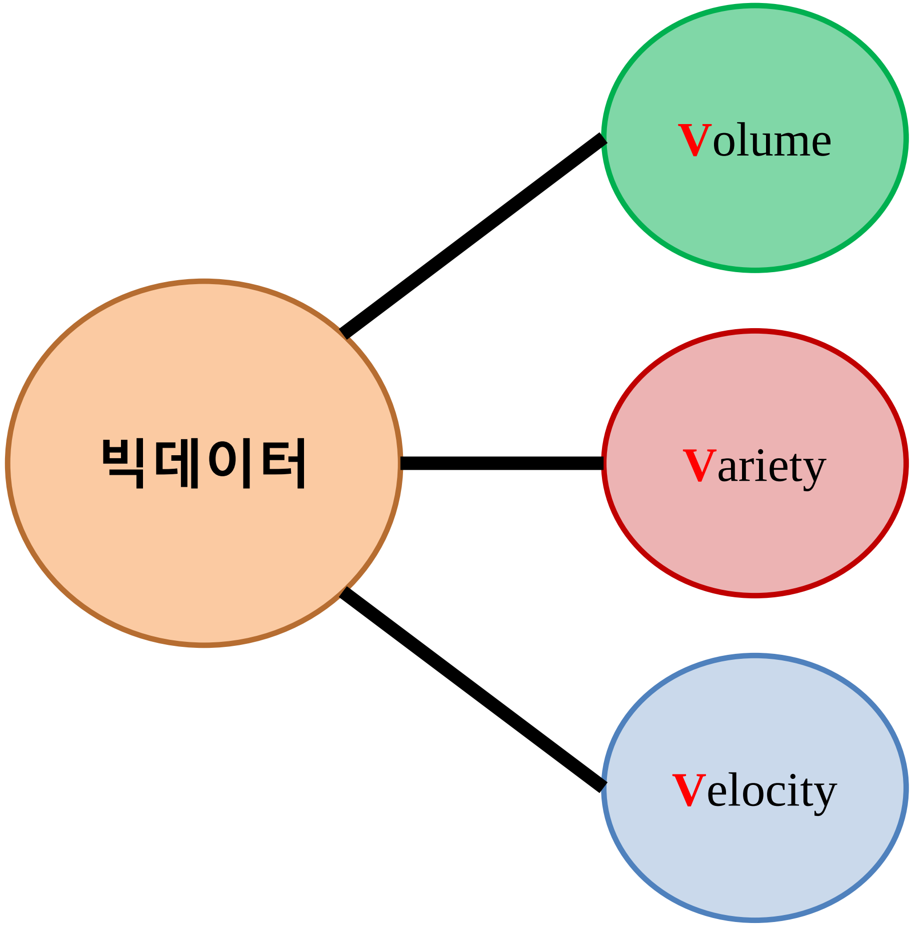
데이터 과학?



→ 데이터 과학은 빅데이터, 머신러닝, AI (인공지능)을 다루는 학문입니다.

빅데이터?

빅데이터의 정의 (3V): Gartner 그룹의 Doug Laney.



→ 데이터의 **양**.

→ 테라 바이트급 이상의 데이터.

→ 데이터 포맷의 **다양성**.

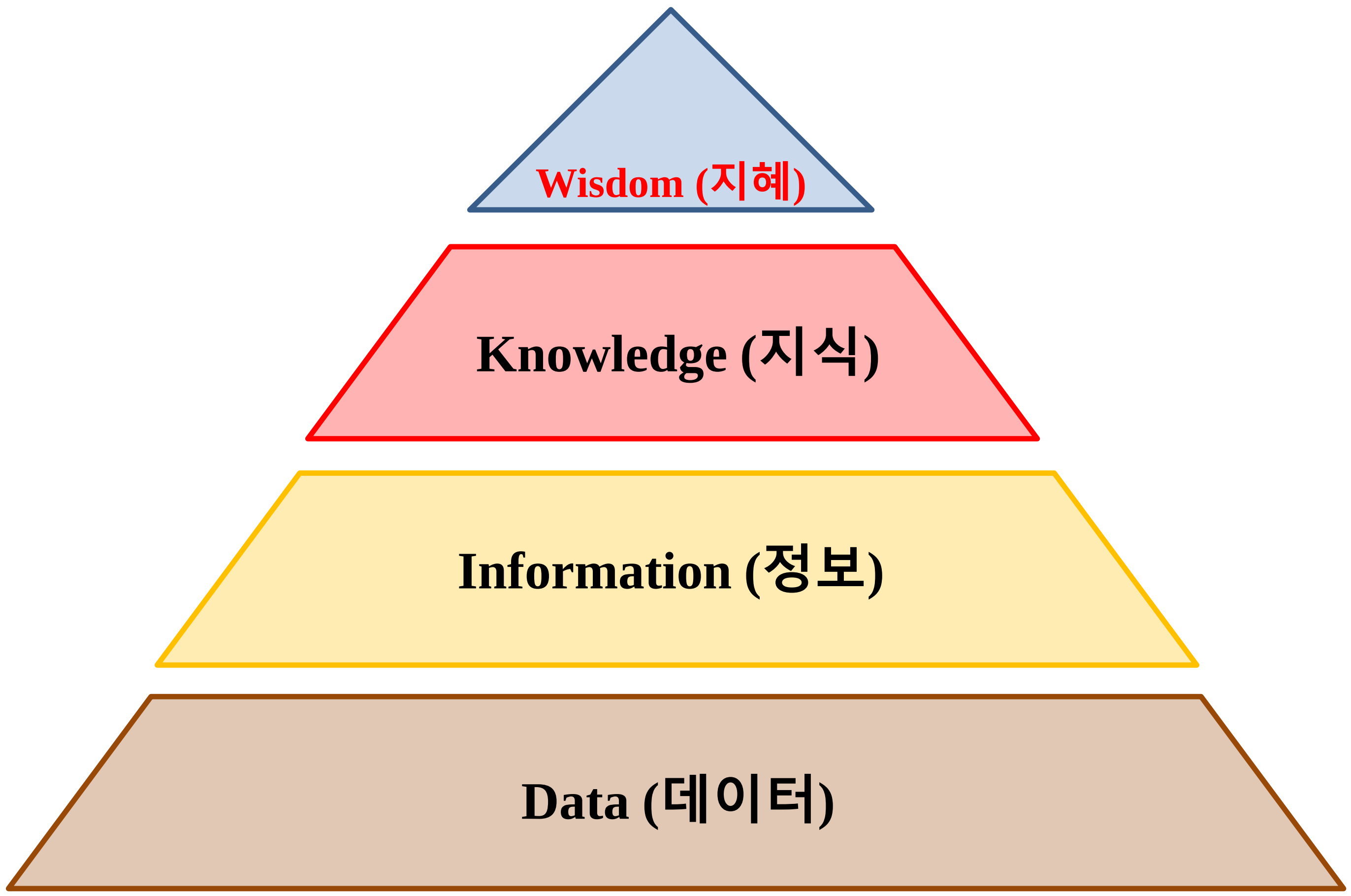
→ 문자, 사진, 동영상, 음향, 등.

→ 신규 데이터의 생성 **속도**.

→ 예) 항공기 센서 데이터 ~ TB/Hour.

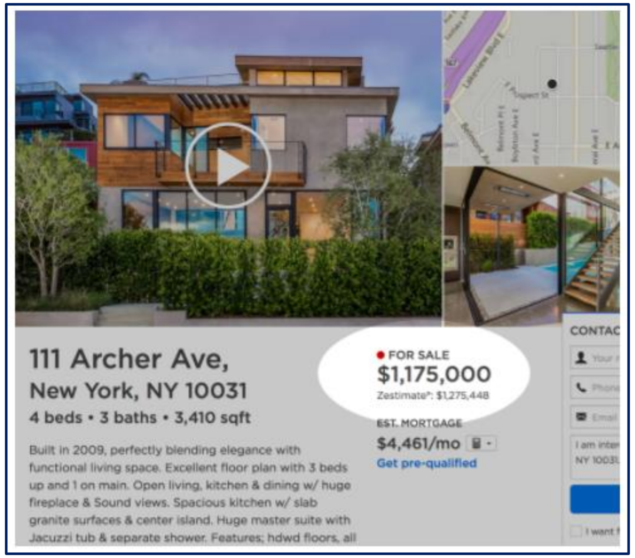
데이터 활용의 궁극적인 목표

데이터와 정보와의 관계: DIKW 피라미드



빅데이터 : 응용 사례

미국 최대의 부동산 가격 감정사이트 Zillow:



→ 1억 1천만 건 이상의 부동산 리스트.

→ 수백여 가지의 변수를 사용하여 가격 감정. 14% → 5% 이하의 오차.

기계학습?

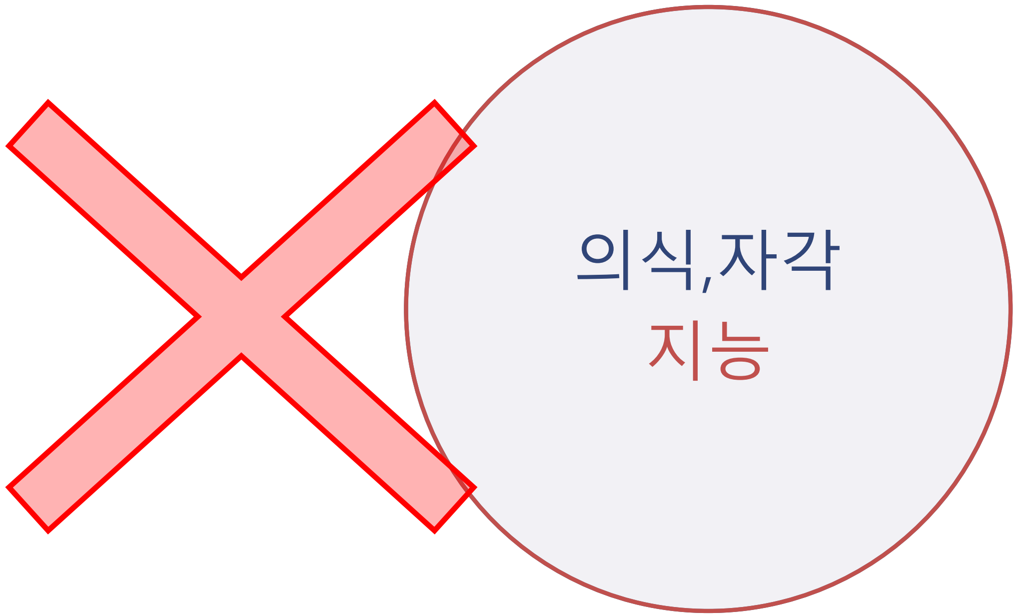
~~러닝머신~~ ???

머신러닝 !
(Machine Learning)

→ **통계 모형**과 데이터를 사용한 학습과 예측을 의미합니다.

인공지능 (AI)?

의식과 지능의 분리:



인공지능

인간

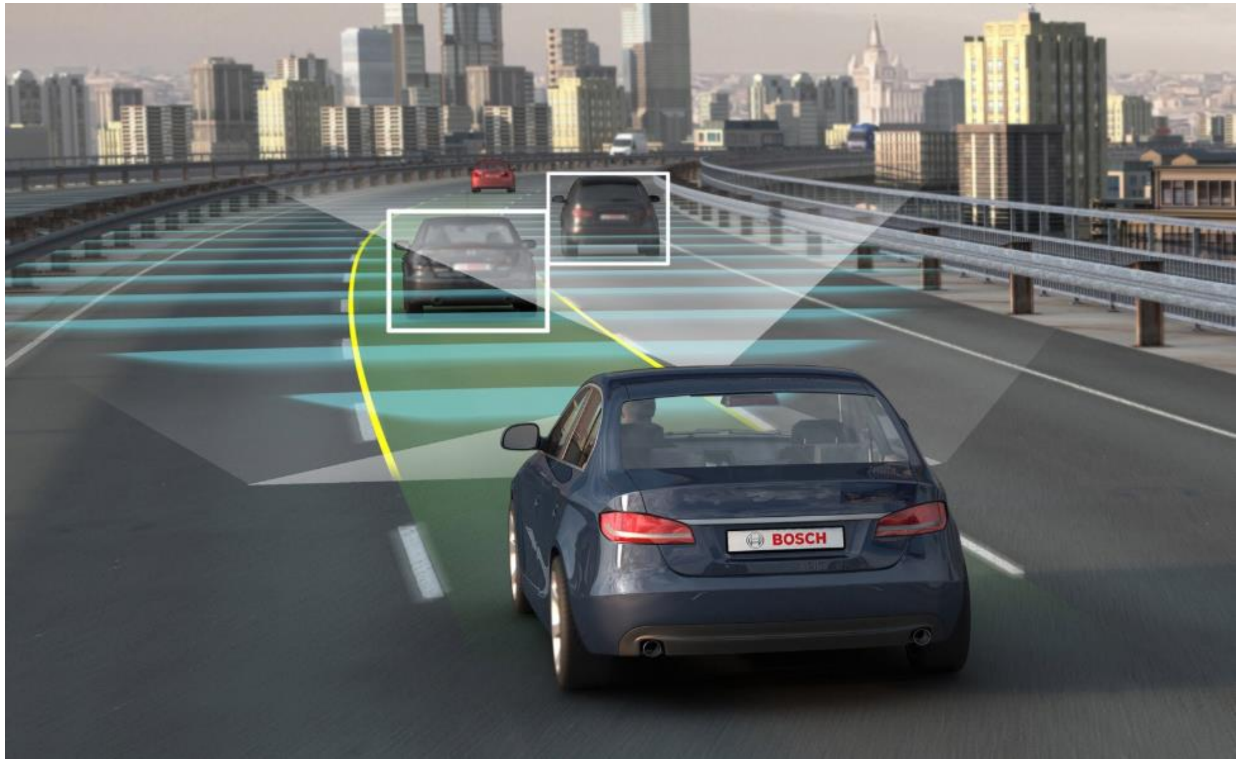
인공지능

??

동물??

고양이??

자율주행 차량:



인공지능 : 응용 분야

통역/번역:

인공지능 : 응용 분야

자산관리:

회의론

AI/빅데이터 **열풍**과 **회의론**:

- "빨리 끓어오른 냄비가 빨리 식는다"라는 일종의 거품현상 우려.
- 성급한 투자에 비해서 미미한 성과.
- 성과의 과대 포장 논란.
- 관련 실무자의 통찰력 도출 능력이 관건.

AI/빅데이터 시대 : 인재상

필요한 인재의 역량:

하드 스킬

(코딩 + 수학)

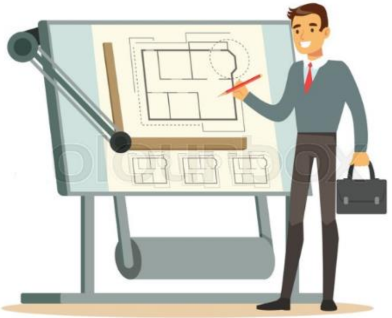
\+

소프트 스킬

(창의력 + 스토리텔링 + 커뮤니케이션)

AI/빅데이터 시대 : 인재의 역량

예술



+창의력

기술

+감성

기능 (스킬)

순서

확률

열공!

열공!

열공!

열공!

열공!

열공!

확률?

확률?

확률?

확률?

확률 기초 : 시행과 사건

시행:

- 같은 조건 아래에서 **반복**할 수 있음.
- 그 결과가 **우연**에 의해서 결정됨.
- 가능한 모든 결과를 알 수 있는 관찰 또는 **실험**.

예). 주사위 던지기.

예). 동전 던지기.

확률 기초 : 시행과 사건

표본공간과 사건:

- 시행으로 가능한 결과 $\Rightarrow$ **사건** (event).
- 시행으로 가능한 모든 결과 $\Rightarrow$ **표본공간** (sample space, $\boldsymbol{S}$).

예). 주사위: $S = \{1, 2, 3, 4, 5, 6\}$

예). 동전: $S = \{T, H\}$

확률 기초 : 시행과 사건

표본공간과 사건: 한개의 주사위 던지기 예 #1

1). 5의 눈이 나오는 사건:

$$E_1 = \{5\}$$

2). 홀수의 눈이 나오는 사건:

$$E_2 = \{1, 3, 5\}$$

3). 3 이상의 눈이 나오는 사건:

$$E_3 = \{3, 4, 5, 6\}$$

확률 기초 : 시행과 사건

표본공간과 사건: 10원짜리 동전 한개와 100원짜리 동전 한개를 동시에 던지기 예

1). 표본공간: H = 앞면, T=뒷면.

$$S = \{(H,H),(H,T),(T,H),(T,T)\}$$

2). 두 개 모두 뒷면이 나오는 사건:

$$E_1 = \{(T,T)\}$$

3). 두 개 모두 같은 면이 나오는 사건:

$$E_2 = \{(H,H),(T,T)\}$$

4). 적어도 한쪽이 앞면이 나오는 사건:

$$E_3 = \{(H,H),(T,H),(H,T)\}$$

확률 기초 : 시행과 사건

표본공간과 사건: 한개의 주사위 던지기 예 #2

1). 짝수의 눈이 나오는 사건 A와 3이상의 눈이 나오는 사건 B:

$$A = \{2, 4, 6\}$$

$$B = \{3, 4, 5, 6\}$$

2). 짝수 이면서 3이상의 눈이 나오는 사건:

$$A \cap B = \{4, 6\}$$

3). 짝수 또는 3이상의 눈이 나오는 사건:

$$A \cup B = \{2, 3, 4, 5, 6\}$$

확률 기초 : 시행과 사건

표본공간과 사건: 한개의 주사위 던지기 예 #2

4). 짝수의 눈이긴 한데 3이상이 아닌 사건:

$$A - B = \{2\}$$

5). 3이상의 눈이긴 한데 짝수가 아닌 사건:

$$B - A = \{3, 5\}$$

6). 이제 홀수의 눈이 나오는 사건을 C라고 정의하면:

$$A = \{2, 4, 6\}$$

$$C = \{1, 3, 5\}$$

$A \cap C = \phi$ “**배반사건**”

확률 기초 : 시행과 사건

표본공간과 사건: 한개의 주사위 던지기 예 #2

7). 3이상의 눈이 아닌 사건:

$$B = \{3, 4, 5, 6\}$$

$$B^c = S - B = \{1, 2\}$$ “**여사건**”

확률 기초 : 정의

수학적 확률의 정의:

- N 이 한 시행에 따라서 일어날 수 있는 모든 경우의 수이고,
- N_A이 사건 A가 일어나는 경우의 수일때 **수학적 확률 *P(A)***는 다음과 같다:

$$P(A) = \frac{N_A}{N}$$

예). 하나의 주사위를 던졌을 때 눈이 짝수일 확률:

$$\text{눈이 짝수일 확률} = \frac{\text{눈이 짝수인 경우의 수}}{\text{모든 경우의 수}} = \frac{3}{6} = \frac{1}{2}$$

확률 기초 : 기본 성질

확률의 기본 성질:

- 임의의 사건 A에 대한 확률은 다음 조건을 충족한다.

$$0 \leq P(A) \leq 1$$

- 근원사건의 확률 p_i를 모두 더하면 1이 되어야 한다.

$$p_1 + p_2 + \cdots + p_N = 1$$

$\Rightarrow$ 즉, 표본공간 S에 대한 확률 $P(S) = 1$ 이다.

확률 기초 : 예제 #0101

두 주사위를 동시에 던질 때, 다음 물음에 답하라:

1). 두 개가 서로 다른 눈이 나올 확률을 구하여라.

$$\rightarrow \text{확률} = \frac{\text{기대하는 것이 일어나는 경우의 수}}{\text{일어날 수 있는 모든 경우의 수}}$$

$$= \frac{36-6}{36} = \frac{30}{36} = \frac{5}{6}$$

2). 두 개의 눈의 합이 7이될 확률을 구하여라.

→ 눈의 합이 7인 경우는 (1,6), (2,5), (3,4), (4,3), (5,2), (6,1)과 같이 6가지 이므로

$$\text{확률} = \frac{6}{36} = \frac{1}{6}$$

확률 기초 : 정의

통계적(경험적) 확률의 정의:

- 어떤 조건 아래에서 실험 또는 관측한 자료의 총수를 N이라 하고, 그 중에서 어떤 사건 A가 일어난 횟수를 N_A라 할 때 **상대도수**는 다음과 같이 정의한다:

$$\text{상대도수} = \frac{N_A}{N}$$

- 그러면 사건 A가 일어날 **통계적 확률**은 다음과 같이 상대도수의 극한이다:

$$\text{통계적 확률} = \lim_{N \to \infty} \frac{N_A}{N}$$

확률 기초 : 예제 #0102

다음 통계적(경험적) 확률에 대한 물음에 답하시오:

1). 통계조사 결과 출생아 1000명 중 출생 후 1년까지 980명이 생존해 있다면, 출생아의 1년 생존확률은?

$$\rightarrow \text{확률} = \frac{\text{기대하는 것이 일어난 횟수}}{\text{실험한 모든 횟수}} = \frac{980}{1000} = 0.98$$

2). 한 개의 주사위를 10000번 던져서 1의 눈이 1650번 나왔다. 1의 눈이 나올 확률을 구하라.

$$\rightarrow \text{확률} = \frac{\text{기대하는 것이 일어난 횟수}}{\text{실험한 모든 횟수}} = \frac{1650}{10000} = 0.165 \approx \frac{1}{6}$$

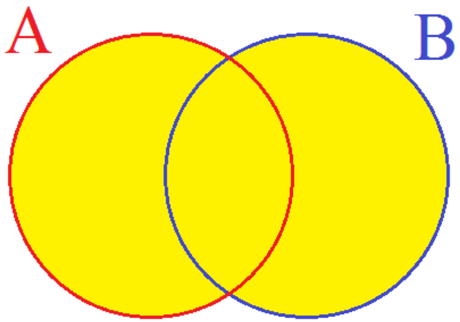
순서

확률의 덧셈 정리

확률의 덧셈 정리:

- 어떤 사건 A **또는** B가 일어날 확률 $P(A \cup B)$은 다음과 같이 계산한다.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



확률의 덧셈 : 예제 #0103

하나의 주사위를 던질 때, 짝수의 눈 또는 3 이상의 눈이 나올 확률은?

→ 짝수의 눈이 나오는 사건 A와 3이상의 눈이 나오는 사건 B:

$$A = \{2, 4, 6\} \quad \Rightarrow \quad P(A) = \frac{3}{6}$$

$$B = \{3, 4, 5, 6\} \quad \Rightarrow \quad P(B) = \frac{4}{6}$$

그리고, $P(A \cap B) = P(\{4, 6\}) = \frac{2}{6}$

→ 확률의 덧셈 정리를 적용하면:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

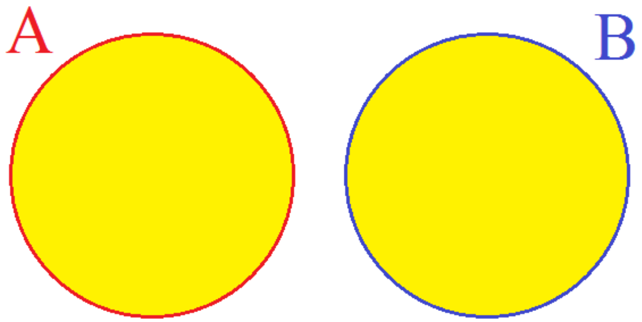
$$= \frac{3}{6} + \frac{4}{6} - \frac{2}{6} = \frac{5}{6}$$

확률의 덧셈 정리

확률의 덧셈 정리:

- 만약에 A와 B가 동시에 일어날 수 없는 "배반사건"인 경우, $P(A \cap B) = 0$ 이며:

$$P(A \cup B) = P(A) + P(B)$$



예). 어떤 사람의 혈액형이 동시에 O형과 A형이 될 수 없다. → "배반사건"

확률의 덧셈 : 예제 #0104

100명의 학생 중 혈액형이 O형, A형, B형, AB형인 학생이 각각 29명, 41명, 18명, 12 명이라고 한다. 이중에서 임의로 한명을 뽑을 때, O형이거나 A형일 확률을 구하여라:

→ 확률의 덧셈 정리를 적용하면:

$$P(O) = \frac{29}{100}$$

$$P(A) = \frac{41}{100}$$

$P(O \cap A) = 0$ → 동시에 A형과 O형이 될 수는 없다. "**배반사건**"

그러므로, $P(O \cup A) = P(O) + P(A)$

$$= \frac{29}{100} + \frac{41}{100} = \frac{7}{10}$$

순서

1. 확률 I:

   1.1. 시대적 배경.

   1.2. 확률의 정의.

   1.3. 확률의 덧셈.

   1.4. 확률의 곱셈.

확률의 곱셈 정리

확률의 곱셈 정리: 조건부 확률 (Conditional Probability)

- 확률이 0이 아닌 사건 A와 B에 대해서 사건 A가 일어났다는 전제로 사건 B가 일어날 확률을 **조건부 확률**이라 부르며 $P(B|A)$와 같이 표기한다. 그리고 다음과 같은 관계가 성립된다.

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

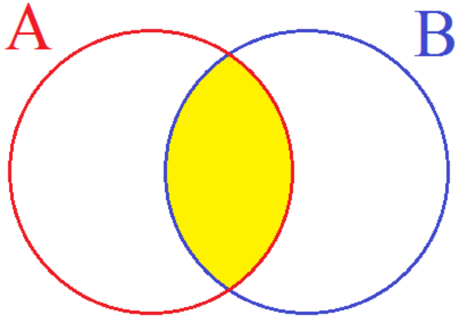
확률의 곱셈 정리

확률의 곱셈 정리: 조건부 확률 (Conditional Probability)

- 이전 수식을 다음 관계로 바꾸어 볼 수 있다:

$$P(A \cap B) = P(B|A)P(A)$$

$$= P(A|B)P(B)$$



확률의 곱셈 정리 : 예제 #0105

주머니 속에 흰 공 4개, 붉은 공 6개가 들어있다. 공을 한 개씩 두 번 꺼낼 때, 다음 각 경우에 대하여 두 개가 모두 흰 공일 확률을 구하여라:

1). 처음에 꺼낸 공을 다시 넣지 않는 경우.

→ 첫 번째, 두 번째에 흰 공이 나오는 사건을 각각 A, B라 한다. 그리고 구하고자 하는 확률은 $P(A \cap B)$이다. 그러면, 확률의 곱셈 정리를 적용해 본다.

$$P(A \cap B) = P(B|A)P(A)$$

$$= \frac{3}{9} \times \frac{4}{10} = \frac{2}{15}$$

확률의 곱셈 정리 : 예제 #0105

주머니 속에 흰 공 4개, 붉은 공 6개가 들어있다. 공을 한 개씩 두 번 꺼낼 때, 다음 각 경우에 대하여 두 개가 모두 흰 공일 확률을 구하여라:

2). 처음에 꺼낸 공을 다시 넣는 경우.

→ 그러면 A와 B는 완전한 독립사건이 된다. 그러므로

$$P(A \cap B) = P(B|A)P(A) = P(B)P(A)$$

$$= \frac{\mathbf{4}}{\mathbf{10}} \times \frac{4}{10} = \frac{4}{25}$$

확률의 곱셈 정리 : 베이즈 정리

베이즈 정리의 도출:

- 확률의 곱셈정리를 이용하여 다음 베이즈 정리를 도출해 낼 수 있다:

$$P(A \cap B) = P(B|A)P(A)$$
$$= P(A|B)P(B)$$

$$\Downarrow$$

$$\boxed{P(A|B)P(B) = P(B|A)P(A)}$$

"베이즈 정리"

확률의 곱셈 정리 : 베이즈 정리

베이즈 정리의 활용:

- 확률의 곱셈정리를 이용하여 다음 베이즈 정리를 도출해 낼 수 있다:

$$P(A|B)P(B) = P(B|A)P(A)$$

- 그러므로 다음이 성립된다.

$\Rightarrow$ 여기서 A^c는 A 이외의 사건을 의미하는 "여사건"이다.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

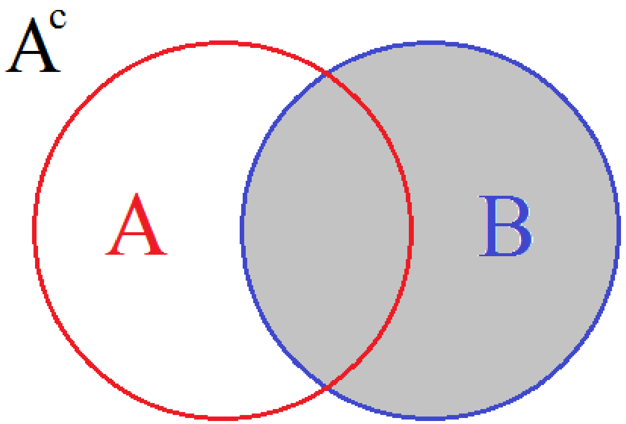
확률의 곱셈 정리 : 베이즈 정리

베이즈 정리의 활용:

- 확률의 곱셈정리를 이용하여 다음 베이즈 정리를 도출해 낼 수 있다:

$$P(A|B) = \frac{P(B|A)P(A)}{\boxed{P(B|A)P(A) + P(B|A^c)P(A^c)}}$$

$P(B)$



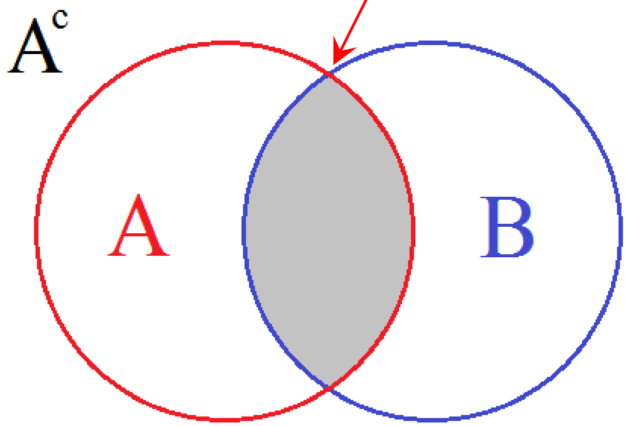
확률의 곱셈 정리 : 베이즈 정리

베이즈 정리의 활용:

- 확률의 곱셈정리를 이용하여 다음 베이즈 정리를 도출해 낼 수 있다:

$$P(A|B) = \frac{P(B|A)P(A)}{\boxed{P(B|A)P(A)} + P(B|A^c)P(A^c)}$$

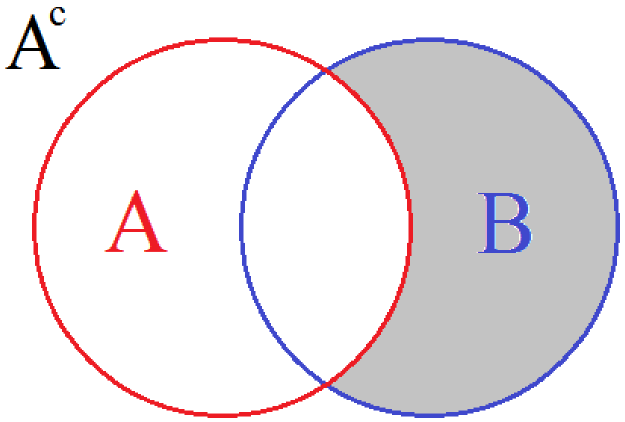
$$P(B \cap A)$$



베이즈 정리의 활용:

- 확률의 곱셈정리를 이용하여 다음 베이즈 정리를 도출해 낼 수 있다:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + \boxed{P(B|A^c)P(A^c)}}$$



$P(B \cap A^c)$

확률의 곱셈 정리 : 예제 #0106a

변종 인플루엔자가 돌고있다. 전체 감염자의 확률은 3%라고 한다. 보건부는 진단 방법을 개발 했는데 실제 감염자 중에서 98%를 정확하게 양성(+)으로 진단하고 또한 실제 비감염자 중에서 95%를 정확하게 음성(-)으로 진단할 수 있다고 한다. 본인도 이 검사를 받아 보았는데 결과는 양성(+)으로 나왔다. 실제 이 변종 인플루엔자에 결렸을 확률은?

→ D가 피검사자가 변종 인플루엔자에 걸릴 사건라면 D^c는 변종 인플루엔자에 걸리지 않을 사건이다. 문제의 조건을 정리해 보면

$P(+|D) = 0.98$ "민감도 (Sensitivity, Recall)"

$P(-|D^c) = 0.95$ "특이도 (Specificity)" $\Rightarrow P(+|D^c) = 1 - P(-|D^c) = 0.05$

$P(D) = 0.03$ "감염률"

문제는 $P(D|+)$를 요구하고 있는 것이다.

확률의 곱셈 정리 : 예제 #0106a

변종 인플루엔자가 돌고있다. 전체 감염자의 확률은 3%라고 한다. 보건부는 진단 방법을 개발 했는데 실제 감염자 중에서 98%를 정확하게 양성(+)으로 진단하고 또한 실제 비감염자 중에서 95%를 정확하게 음성(-)으로 진단할 수 있다고 한다. 본인도 이 검사를 받아 보았는데 결과는 양성(+)으로 나왔다. 실제 이 변종 인플루엔자에 걸렸을 확률은?

→ 그러면 베이즈 정리를 적용해 본다.

$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D)+P(+|D^c)P(D^c)}$$

$$= \frac{0.98 \times 0.03}{0.98 \times 0.03 + 0.05 \times 0.97} \cong \mathbf{0.377}$$

확률의 곱셈 정리 : 예제 #0106b

100 개의 관측치 중에서 6 개의 경우 반응 변수 (Y)의 실제 값이 1이며,

나머지 94 경우에는 0이다. 어느 머신러닝 모형이 민감도 = 0.92, 특이도 = 0.90와

같은 성능을 보인다고 가정한다. 새롭게 설명 변수의 관측값 (X_{test})을 사용하여

예측한 반응 변수의 값이 1이다.

어느 정도의 확률로 이 예측이 정확한가?

→ 이미 알고 있는 것: $P(1) = 0.06$, $P(0) = 0.94$

$P(Predicted\ 1|Actual\ 1) = 0.92$ ⇐ "민감도"

$P(Predicted\ 0|Actual\ 0) = 0.90$ ⇐ "특이도"

또한 $P(Predicted\ 1|Actual\ 0) = 1 - P(Predicted\ 0|Actual\ 0) = 0.1$ 이다.

확률의 곱셈 정리 : 예제 #0106b

100 개의 관측치 중에서 6 개의 경우 반응 변수 (Y)의 실제 값이 1이며,

나머지 94 경우에는 0이다. 어느 머신러닝 모형이 민감도 = 0.92, 특이도 = 0.90와

같은 성능을 보인다고 가정한다. 새롭게 설명 변수의 관측값 (X_{test})을 사용하여

예측한 반응 변수의 값이 1이다.

어느 정도의 확률로 이 예측이 정확한가?

→ 그러면 정답은 다음과 같다:

$$P(Actual\ 1|Predicted\ 1) = \frac{P(Predicted\ 1|Actual\ 1)P(1)}{P(Predicted\ 1|Actual\ 1)P(1) + P(Predicted\ 1|Actual\ 0)P(0)}$$

$$= \frac{0.92 \times 0.06}{0.92 \times 0.06 + 0.1 \times 0.94} \cong 0.37$$

모듈 #1 : 끝

문의:

sychang1@gmail.com